MISB MOTION IMAGERY STANDARDS BOARD **Technical Reference Material** **H.264 Study for Mobile Devices**	**MISB TRM 1012** **30 Sep 2010**

1 Scope

Delivery of H.264 compressed Full Motion Video (FMV) to the tactical edge is becoming more prevalent in theatre with mobile devices such as ROVER5, and in the near future iPhone/Android footprint devices. Important to maximizing the utility of these mobile devices is tailoring the source FMV to the capabilities of the device. Power consumption, display resolution, CPU processing power, and the bandwidth-constricted networks the data must travel all factor into the resulting user experience—and thus usefulness of the device. Examining H.264 parameters while considering these factors will help system developers optimize the choice of FMV spatial and temporal resolution, and help to identify those parameters that may cause the video to display improperly.

FMV to the tactical edge—considered as that to a disadvantaged user— is generally labeled as providing Situational Awareness (SA) FMV. SA FMV can be of various data rates, levels of quality, and provided using different delivery vehicles to meet the constraints of channel bandwidth, channel type (wired/wireless), and client device. Understanding of the H.264 compression profiles, structures and features across this wide variation in usage should prove valuable guidance.

While very important to overall system performance, topics of network considerations, such as network type (for example 3G, 4G, other), and parsing metadata from the video stream are outside the scope of this study.

2 Mobile Devices

Three mobile devices were evaluated: the Apple iPhone3, Android, and ROVER5.

3 H.264 Profile

The H.264 baseline profile was the profile tested across the mobile devices since this is what they support. A 1280x720p file served as the reference test file for performance evaluations. The content within this file includes aerial views of pans across a plane in flight, trucks moving along dirt roads, and people waving on the ground.

Two types of tests were done:

1. Assessment of compatibility with H.264 Levels 1 through 3 (Table 1)
2. Assessment of specific H.264 coding options

4 Test Scenario 1

The reference file was first transcoded according to the constraints of a particular *level* as specified within H.264, such as picture size, frame rate and data rate as shown in Table 1. All three mobile devices demonstrated that they could properly decode the transcoded reference material.

H.264 Level	Picture Size	Frame Rate (maximum)	Compressed Bit Rate (maximum)
1	QCIF	15	74 kbps
1b	QCIF	15	128 kbps
1.1	CIF or QCIF (176x144)	7.5 (CIF) / 30 (QCIF)	192 kbps
1.2	CIF	15	384 kbps
1.3	CIF	30	768 kbps
2	CIF (352x288)	30	2 Mbps
2.1	HHR (352x480)	30 / 25	4 Mbps
2.2	SD (720x480)	15	4 Mbps
3	SD	30 / 25	10 Mbps

Table 1 – H.264 levels tested

5 Test Scenario 2

In this series of tests, the reference file was transcoded for various data rates, GOP (Group of Pictures) structure and size, temporal frame rates, motion estimation method, motion search area and picture size. The coding method is CAVLC.

5.1 *Tested Parameters*

- GOP sizes between 15 and 120 and infinite GOP
- Frame rates from 1-60 Hz
- Motion estimation: quarter and half pixel
- Search area: 8x8 and16x16
- Aspect ratio: 4x3 and 16x9

5.1.1 GOP Size

There was no noticeable difference in subjective quality of the video between GOP sizes in the range of 15 to 120. This is principally because the differences—while observable on a larger screen—are not perceptible on the small viewing screens. An infinite size GOP did produce poor quality at the low bit rates tested. A GOP size of 120 produced a better quality image—although at the expense of a slight increase in bit rate.

Long GOP sizes are desirable because they require less decoder computation, and together with the lower bit rate decrease the strain on the battery thereby prolonging device usability. On the other hand, a long GOP sequence contains fewer I-frames and is thus less robust susceptible to packet loss. Therefore, the desired quality of the video must be traded against the length of device usability. A further consequence of a long GOP is that the decoder must wait until an Iframe is received to begin decoding, which impacts how quickly a client can begin decoding and present the video for viewing.

5.1.2 Frame Rate

All three devices were tested using the transcoded reference file at frame rates from 1 Hz to 60 Hz. Android played all frame rates tested, although the video was not smooth at the 1 to 2 Hz rates. The iPhone and ROVER5 played content up through 30 Hz. The Rover5 lower limit is 2 Hz, while the iPhone is 1 Hz. Again, the playback at these low frame rates was not as smooth as at higher rates.

5.1.3 Motion Estimation

Two methods for motion estimation were evaluated: half pixel and quarter pixel. These methods represent the fineness that a determination of movement of objects within a video sequence is made. Quarter-pixel motion estimation requires far greater computation than half-pixel estimation, but improves object movement prediction immensely. Since these decisions are made in the encoder, the impact of motion estimation in a decoder is principally one of how accurate the motion is portrayed in the video upon decoding and rendering.

The reference sequence was transcoded using both types of motion estimation. Subjectively, there was no difference in the portrayal of movement across the three devices. This result can be attributed to two potential reasons: one, with such a small display it is very difficult to notice any difference; two, the structure in the content was insensitive to the motion estimation technique used. That is, content with finer object movement may produce noticeable differences. However, since the content is typical of that taken from an airborne platform, similar results to those here are anticipated.

Since encoded video with the quarter-pixel motion estimator will be of higher quality and the subsequent produced data rate likely lower, it is recommended that the quarter-pixel motion estimation be used for encoding when possible.

5.1.4 Search Area

Similar to motion estimation, a larger area covered in searching for objects in the encode process will allow for better motion tracking of objects that move greater distances between frames. Better quality video is produced in such cases when using a larger search area. Search areas of 8x8 and 16x16 pixels were tested. There was no noticeable difference across the three mobile platforms. As discussed above, this can be likewise attributed to the same two possible reasons: small display size, or insensitive content structure. In general, since more efficient coding can be produced with a wider search area producing lower data rates it is recommended that this be done.

5.1.5 Aspect Ratio

The aspect ratio for standard definition (SD) video is 4x3, while for high definition (HD) content the aspect ratio is 16x9. This corresponds typically to a 640x480 pixel density for SD content and a 1280x720 pixel density for HD content. The ROVER5 has a SD display size of 640x480, while the iPhone and Android exhibit fewer pixels at a display format of 480x320.

The reference test file was transcoded into 720x480 pixels. All three devices displayed the content without issue. The iPhone and Android each scaled the content in size down to meet the 480x320 display format. It was not clear if the ROVER5 clipped off 40 pixels on each horizontal edge to meet its display format of 640 pixels, if it squeezed the video horizontally to fit, or if it scaled both dimensions to maintain the picture aspect ratio (which would create black bands at the vertical top and bottom).

For pixel densities other than the native display format, black bars at the top/bottom or left/right sides are visible.

	iPhone 3	**Android**	**ROVER5**
File Format(s) (supported)	.m4v .mp4 .mov	.3gp .mp4 .3g2	MPEG-2 Transport Stream
Display Resolution (maximum)	480x320	480x320	640x480
Operating System	iOS 3.0	1.5 (Cupcake)	Monta Vista Linux
Bit Rate (maximum)	12 Mbps	12 Mbps	6 Mbps
Frame Rate (range)	1-30	1-60	2-30
Streaming Protocol	Adaptive	RTSP	MPEG-2 TS
Battery Life (hours)	6*	5.5*	10 (w-battery pack)
*Continuous video playback in nominal condition			

Table 2 – Specifications of Mobile Devices Tested

6 Summary

This study addressed some of the current capabilities of mobile handheld devices available as of November 2009. Mobile handhelds continue to improve in battery life, processing power, display resolution, and the ability to decode higher H.264 profiles and levels. Very useful to know is how well these devices operate in streaming the content. Exercising the resiliency tools within H.264 in this light would also prove valuable. Evaluating the option to display metadata through on-screen display (available for ROVER5) or via other means would be an important follow on to this initial study.